CLOCK GATING TECHNIQUE FOR POWER REDUCTION IN DIGITAL DESIGN

by

KHOR PENG LIM

Thesis submitted in fulfillment of the requirements for the degree of Master of Science

December 2012

ACKNOWLEDGEMENT

I would like to express my most sincere thanks to Associate Professor Dr. Mohd Fadzil bin Ain, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, my supervisor, for the encouragement, personal guidance, assistance and valuable suggestions enabling me to steer my research work efficiently and effectively. His wide knowledge in the field of electronic and engineering for research work had been extremely useful for my research work and provided excellent basis for my thesis at the end.

I am very grateful to my co-supervisor, Mr. Lock Choon Hou, Penang Design Centre, Intel Malaysia, for his detailed and constructive comments, and support throughout my work.

Most importantly my heartfelt sincere thanks should go to my beloved parents who have been instrumental in raising me up to the heights that I am in at present with their love, courage and support. I dedicate this thesis to them. My special appreciation and gratitude goes to my brother, and their families for their love and kindness.

I would like to extend my gratitude to the project architect, Mr. Sarwar Zeeshan for providing me an industrial research test case.

My sincere gratitude and thanks in no uncertain terms is expressed herein to my dear colleagues Mr. J-Me, Ms. Diana Tan, Ms. Liew, Mr. Teoh and Mr. Zainal for their invaluable support extended to me in numerous ways during my studies.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

IC	Integrated Circuit
TDP	Thermal Design Power
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
OPC	Optical Proximity Correction
SOC	System-On-Chip
CMOS	Complementary Metal Oxide Semiconductor
DVFS	Dynamic Voltage and Frequency Scaling
DC	Direct Current
PMOS	P-type Metal Oxide Semiconductor
NMOS	N-type Metal Oxide Semiconductor
GPU	Graphic Processor Unit
USB	Universal Serial Bus
RTL	Register Transfer Level
VCD	Value Change Dump
SAIF	Switching Activity Information File

PE	Protocol Engine
BFM	Bus Functional Module
HKMG	High-K dielectric Metal Gate
VCS	Verilog Compiler Simulation
GUI	Graphical User Interface
HVT	High Threshold Voltage
LVT	Low Threshold Voltage
Multi-VT	Multiple Threshold Voltage

TEKNIK PENGGETAN JAM UNTUK PENGURANGAN KUASA DALAM REKABENTUK DIGIT

ABSTRAK

Teknik pengurangan kuasa menjadi unsur yang semakin penting bagi litar digital bersepadu berskala sub-mikron. Teknik-teknik pengurangan kuasa digunakan untuk mengawal penggunaan kuasa litar bersepadu yang beroperasi pada frekuensi yang tinggi. Teknik pengurangan kuasa yang sama tidak semestinya memberi kecekapan yang sama apabila frekuensi litar bersepadu tersebut berubah. Dalam penyelidikan ini, teknik pengurangan kuasa yang dipilih telah diuji dengan litar bersepadu yang beroperasi dalam pelbagai frekuensi. Ini adalah untuk mengkaji kecekapan teknik-teknik pengurangan kuasa apabila litar bersepadu tersebut beroperasi pada frekuensi yang tinggi. Kecekapan teknik pengurangan kuasa menurun menurut proses rekabentuk litar bersepadu. Bagi pusat rekabentuk litar bersepadu yang tidak mempunyai kilang pengeluar wafer litar bersepadu, teknik pengurangan kuasa yang boleh diguna adalah terhad. Penyelidikan ini menumpu kepada teknik penggetan jam supaya memberikan manfaat yang bagi pusat rekabentuk yang berkenaan. Teknologi proses juga merupakan perkara penting untuk memilih jenis teknik pengurangan kuasa yang digunakan. Bagi teknologi proses maju yang melebihi 90nm, kuasa kebocoran menjadi kuasa penggunaan utama bagi litar bersepadu. Dengan itu, penambahan litar untuk mengurangkan kuasa dinamik mungkin memberi kesan yang negatif. Teknologi proses yang digunakan dalam penyelidikan ini termasuk 32nm, 90nm, 350nm dan 500nm. Keputusan penyelidikan ini menunjukkan kecekapan yang sangat positif apabila beroperasi pada frekuensi tinggi tetapi kecekapannya menurun apabila frekuensi operasi menurun. Teknologi proses baru juga

menyebabkan teknik penggetan jam kurang berkesan. Ini disebabkan oleh teknik penggetan jam menumpu untuk mengurangkan kuasa dinamik manakala kuasa kebocoran merupakan kuasa utama yang digunakan dalam teknologi proses baru.

CLOCK GATING TECHNIQUE FOR POWER REDUCTION IN DIGITAL DESIGN

ABSTRACT

Power reduction techniques become increasingly important to the deep sub-micron scale digital integrated circuit (IC) design. Multiple power reduction techniques are used to keep the power consumption under control even when the operating frequency is high. Same power reduction technique might not give the same power saving efficiency when the operating frequency increases. Power reduction effectiveness decreases follows downward of the design flow. For an IC design house without fabrication factory, levels of power optimization in the design flow are very limited. In this research, selected power reduction techniques are used with different operating frequency to investigate the effectiveness of the techniques in a high speed design. This research focused on the clock-gating power convergence technique to bring the power optimization benefit for the IC design houses that without fabrication factory. With the same power reduction technique, different implementation of the technique will give different efficiency. This research included different approach of clock-gating in a few scenarios to investigate the real world situation. Process technology plays the important role in selecting power convergence techniques to be implemented. With advance process technology below 90nm scale, the leakage power consumption becomes dominant. Hence, adding additional logic to reduce dynamic power consumption might give worse result. This research included few technology libraries which are 32nm, 90nm, 350nm and 500nm for comparison. The result shows that clock-gating technique is very efficient at high speed operating frequency but the benefit decreases when running in low operating

frequency. New process technologies also shows that clock-gating technique is not so efficient due to the transistor device is leakage power dominant while clock-gating is focusing on reducing dynamic power consumption.

CHAPTER 1

INTRODUCTION AND OBJECTIVE

1.1. Introduction

Power convergence technique is a necessary ingredient to design a modern IC. The idea of the power convergence technique is to converge the power profile of the design to meet the desired specification. The power convergence techniques are essentially the power reduction techniques apply from the beginning of the IC design flow (Architectural) to the backend flow of the design cycle (layout). In a high speed design, implementing the techniques requiring much more effort due to the contradiction of speed and power. Both power and timing convergence have to be properly evaluated to balance the trade-off between those two.

There are many power reduction techniques surfaced since the introduction of electronic mobile device. The desktop segment quickly follows when the operation speed approaching Giga-Hertz (GHz) range. There are two major power usage categories by electronic device which are dynamic power and static power consumption. While during the 100nm and above process technology, most of the techniques are focusing on reducing dynamic power consumption. The trend starts to change after entering the deep sub-micron process technology. A big percentage of the total power consumption is taken by the leakage power consumption. These make the implementation of power convergence techniques becomes even more complicated when other factors are being compromised. The trade-off between the techniques, with

the speed and area play a major role in today's IC design. The trade-off depends heavily on applications, available resources and process technology library.

Power consumption is always correlated to energy usage. Generally, power consumption plays an important role in a design's thermal design power (TDP) while energy usage usually tied to the efficiency of the design. TDP is important to decide the cooling and power delivery method for the design especially in mobile sector. Higher power consumption usually leads to higher energy usage but this is not entirely true. Figure 1.1: (a) and (b) shows two task models with the same design. Assuming the frequency and task are the same for both models, model A will gives twice the power consumption compared to model B but both models consume the same amount of energy. The benefit of model A is the time to complete the task 1 has been shorten by half (clock cycle 3 versus clock cycle 6). One of the possible techniques to give the following result will be simply lowering the design operating frequency which will make the performance suffer and the total energy usage is still unchanged.

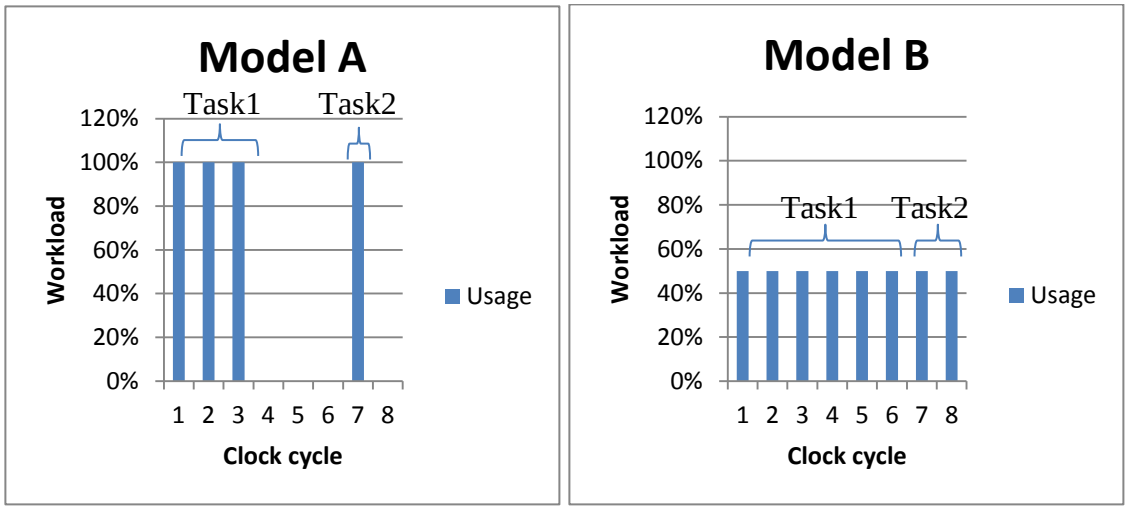

Figure 1.1: (a) and (b) Energy for Two Task Models

When the process technology migration happened, the size of the transistor shrunk. The channel length of the transistor shrinks together with the transistor size which provides faster switching frequency for the transistor. However, the dynamic power consumption of a transistor is a function of switching frequency. Higher switching frequency will increases the dynamic power consumption of the design linearly. Drastically increase the switching frequency to achieve better performance can no longer applied due to unbearable increases of power dissipation. Shorter transistor channel length also leads to lower threshold voltage of a MOSFET. Lower threshold voltage means the MOSFET is leakier and hence, high leakage power is observed on deep sub-micron digital IC design. Leakage power becomes a bigger problem when the transistor density also increases with smaller transistor size. According to the Moore's law, where the number of transistor per unit area will be double every 18 to 24 months.

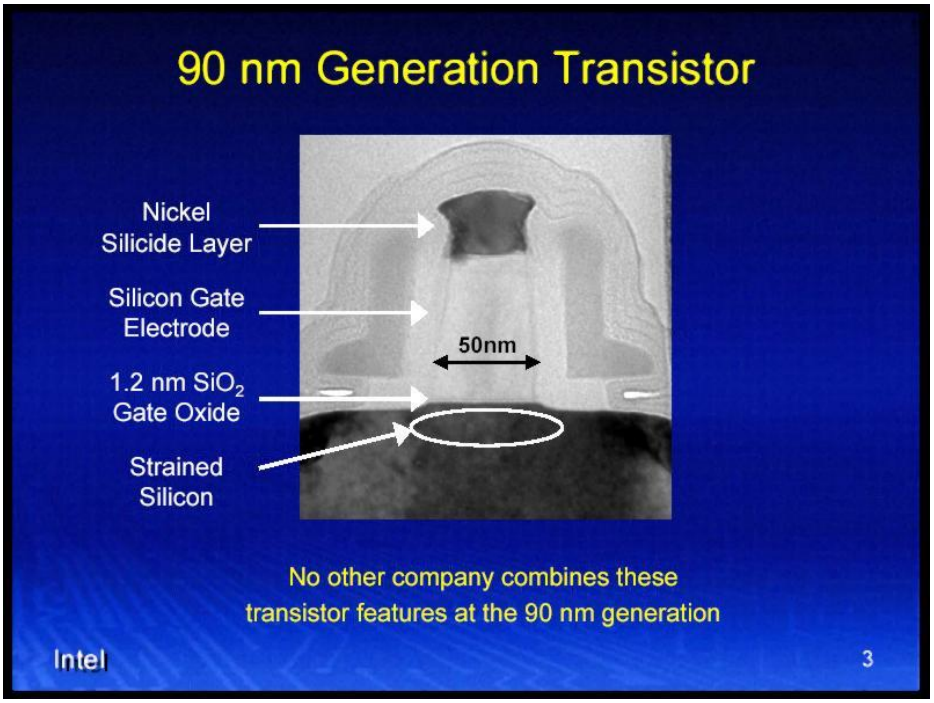

Figure 1.2: 90nm Process Technology (Intel Corp. 2004)

Figure 1.2 shows a transistor fabricated by using Intel® 90nm process technology. The actual physical channel length of the transistor is shorter than the process layout mask due to optical proximity correction (OPC) is in used. Figure 1.3 shows that the transistor count is following Moore's Law closely.

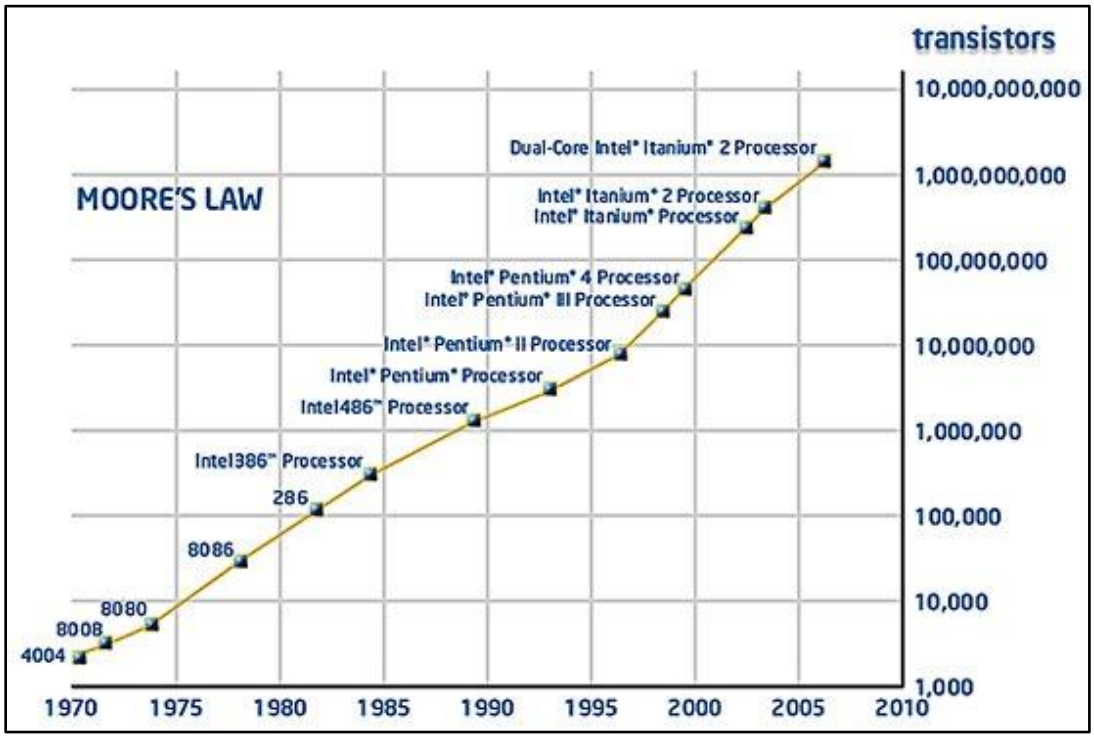

Figure 1.3: Transistor Densities (Intel Corp. 2011)

Power density follows the trend of the transistor density which is unsustainable in a long run. Figure 1.4 shows the prediction of power density if there is no solution for the coming power requirement.

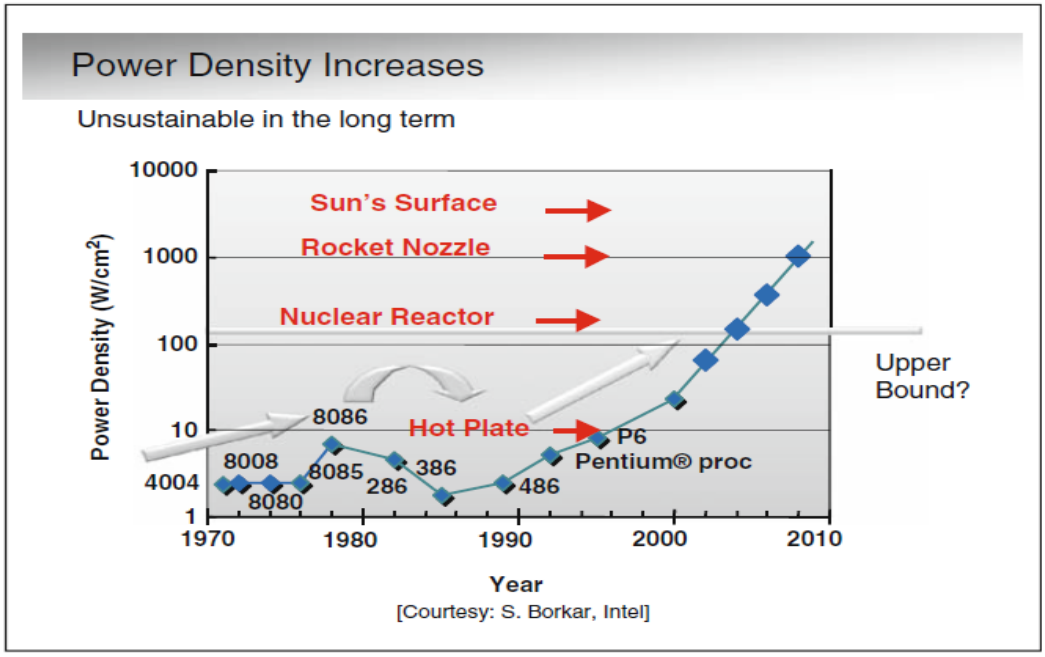

Figure 1.4: Power Densities (Jan M. Rabaey, 2009)

Typical high speed designs are working in Giga-Hertz range. For a modern microprocessor design, 130W-150W of power dissipation is close to the ceiling of the power requirement. To achieve higher operating frequency with the same power envelop, power convergence techniques are required. However, due to the trend of coming technology, same power convergence technique might not be able to be implemented effectively. One of the major factors is changes in process technology. When the transistors are getting smaller, the leakage power starts to dominate the total power consumption of the design. Those power reduction techniques that focusing on reducing the dynamic power consumption may eventually render ineffective or even worsen if the design is running on a slower operating frequency. There are multiple applications that require the power convergence techniques to dynamically switch its focus between reduction of dynamic and leakage power. One of the examples will be System-On-Chip

(SOC) for mobile platform. When the mobile device is in standby mode, there are very limited of switching activities happened. In this case, power reduction on focusing leakage power should be applied. On the other hand, while the mobile device is in active state (surfing internet, playing video or audio), dynamic power starts to take over the majority of the device power consumption. Reducing the dynamic power becomes primary focus in this scenario.

1.2. Problem Statement

For a fab-less IC design houses, they are not able to implement some of the power convergence techniques which involved the transistor level in the technology library. This may not be a critical issue since the effectiveness of power convergence techniques decrease throughout the IC design flow from algorithm downward to the structural layout. Hence, the power convergence techniques chosen to be implemented during upper level of design flow are very crucial. Investigating into various type of power reduction techniques in architectural level will enable the fab-less IC design houses to allocate proper resources and focus into the design.

Modern designs are running in variable frequency to achieve better performance in active mode while reducing power usage in idle or standby mode. Certain techniques might not be suitable for low frequency while others might show negative impact in high frequency.

1.3. Research Objective

The primary research objective is to investigate the efficiency of various types of power convergence techniques with different operating frequencies. Other than frequency, different technology library also will be used to compare the effect of process technology toward the efficiency.

Main focus of the research:

1. Review and identify different type of power convergence techniques available.

2. Investigate the power convergence techniques on different operating frequencies.

3. Establish the best power convergence technique available at time of writing.

1.4. Scope of Research

This research will be focusing on pre-layout power convergence techniques mainly in clock-gating and process technology changes. Pre-layout analysis is more suitable for a design which synthesized using multiple different libraries. Some of the free libraries are missing layout information as well. The modified design will be tested to run in difference operating frequencies from 60MHz to 1GHz.

1.5. Organization of the Thesis

There are five main chapters in the research. The five chapters are organized as introduction, literature review, methodology, result and conclusion. All the working scripts are shown in section APPENDIX.

Chapter 1 describes the overview of the low power techniques and the importance of the techniques. Some background studies in brief with the problem statement also discussed in the chapter. Objectives and scopes of the study are discussed as well.

Chapter 2 states the literature review of the most commonly used power convergence techniques available in the market. Some advance techniques also will be explained in the chapter in brief.

Chapter 3 explains the methodology of the research. The flow of the research including the scripting is covered under this chapter. Libraries, test subject and techniques to be implemented are discussed in this chapter.

Results analysis will be done in chapter 4. The actual results will be compared to theoretical values to investigate the correctness of the result. The different techniques efficiency will be compared and the results will be concluded.

The last chapter will be the conclusion of the research. The research will be concluded based on the coverage of the research objective. Future improvement for the research will be discussed in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1. Power Convergence Techniques

There are many power convergence techniques had been implemented into current microprocessor design. There are three major categories of power reduction area which has been discussed in Section 1.1 Introduction. A deeper analysis on the CMOS power consumption has to be done in order to implement the power convergence techniques. The total power consumption of a MOSFET (Jan Rabaey, 2009) is:

$$P_T = Dynamic\ Power + Short\ Circuit\ Power + Leakage\ Power \quad [2.1]$$

To reduce the dynamic power and short circuit power consumption, the most commonly used methods are clock-gating, data-gating and, dynamic voltage and frequency scaling (DVFS). There are more ways to reduce leakage power, which includes power-gating, multi-threshold voltages, substrate biasing, and multi supply voltage. (S. C. Prasad and K. Roy, 1994)

2.2. Dynamic Power and Short Circuit Power Reduction Techniques

Dynamic power dissipated when the MOSFET charging and discharging internal cell and load capacitance (Veendrick and Harry J. M., 2008). This happened when the logic changes from 1 to 0 or vice versa. Figure 2.1 shows the capacitance model of an inverter.

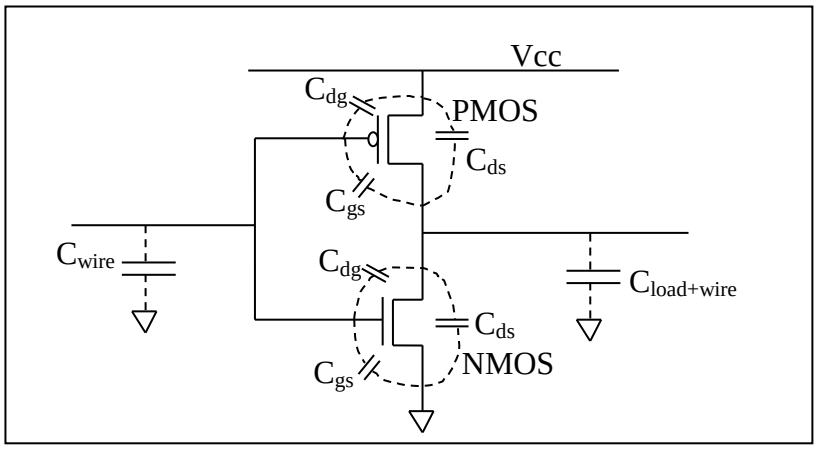

Figure 2.1: Capacitance model of an Inverter

Clock-gating is the most common technique to reduce the dynamic power consumption. Clock-gating works by gating the propagating clock tree if the specific output from a flip-flop is not needed. Figure 2.2 shows the clock-gated flip-flop where the clock is gated by Enable signal.

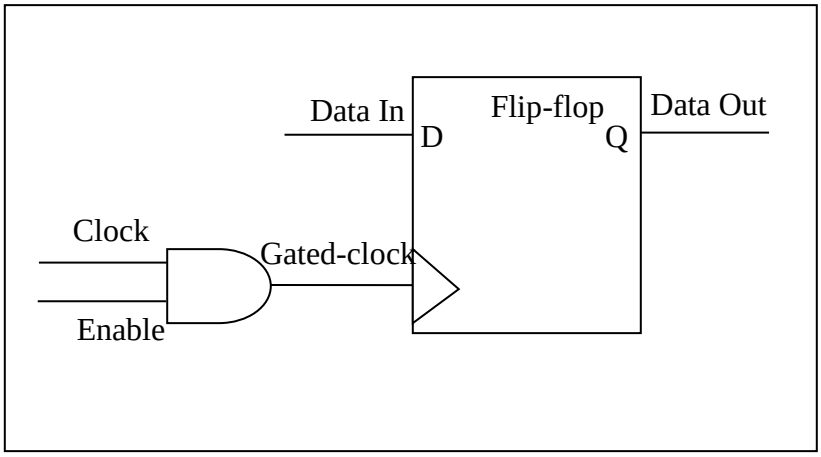

Figure 2.2: Simple Clock-gating Flip-flop

There are two types of clock-gating technique which are coarse grain and fine grain. Coarse grain clock gating is usually used in higher hierarchical level where a single clock gating logic gating the main clock propagates into the block. Fine grain clock gating is normally use in cell level where by each flip-flop's clock is individually gated by its own condition.

To investigate the effect of clock-gating technique towards the dynamic power consumption, the relationship between frequency and the dynamic power consumption are being studied. The dynamic power of a MOSFET digital circuit (Jan Rabaey, 2009) is given by:

$$P_D = \eta C_{eff} V^2 feq \qquad [2.2]$$

where η = switching activity factor, C_{eff} = effective capacitance, V = supply voltage of transistor, feq = operating frequency.

For a simple design, the effective capacitance, C_{eff} will never changes since the physical paths are remain the same. For the purpose of investigation, the supply voltage V is kept constant. The dynamic power equation can be simplified to:

$$P_D \propto \eta feq \quad [2.3]$$

This clearly shows that the switching activity factor, η and the operating frequency feq is directly proportional to the dynamic power consumption of MOSFET. For a clock network without clock-gating implemented, the activity factor is 1. In theory, with a percentage increase of frequency, the same amount of percentage increased can be seen on the dynamic power consumption. Due to limitation of the tools, the short circuit power P_{sc} has to be estimated to 10% of the dynamic power consumption. Equation 2.4 shows the equation to calculate short circuit power based on typical usage model.

$$P_{sc} \approx P_D \times 10\% \quad [2.4]$$

With the respective assumptions, the total power consumption when the DVFS is implemented can be calculated. Referring to Equation 2.5:

$P_l = VI_{leak}$ and the changes of I_{leak} are minimum so

$$P_l \propto V \quad [2.5]$$

When the design has been clock-gated to an activity factor of 0.5, the operating frequency can be increased twice higher within the same power envelop. This will eventually leads to double the performance of the design provided there is no timing violation. Table 2.1 shows the effect of the activity factor towards the dynamic power consumption.

Table 2.1: Percentage of Power Reduction with Different values of frequency and Activity Factor

P(delta)	η	feq (delta)
+50%	1	+50%
+25%	0.5	+50%
+50%	0.5	+100%

Short circuit power consumption happened during the moment when transistors are switching and the supply voltage is directly connected to ground in a short time. The short circuit power consumption (Jan Rabaey, 2009) is given by:

$$P_{sc} = \eta \Delta t I_{sc} V feq \quad [2.6]$$

where η = switching activity factor, Δt = momentary period of short circuit happened, I_{sc}= short circuit current, V = supply voltage of transistor, feq = operating frequency. The same relationship can be seen as Equation 2.6 where the short circuit power consumption is also directly proportional to increase of operating frequency and

switching activity factor. This means that clock-gating can be used to effectively reduce short circuit power as well.

Data-gating is using the same idea of clock-gating except that the gated targets are data lines or bus lines. Figure 2.3 shows a simple data-gated flip-flop where Selector can choose to inject the output of Flip-flop 1 or inject with 0 if the output of Flip-flop 1 is not needed.

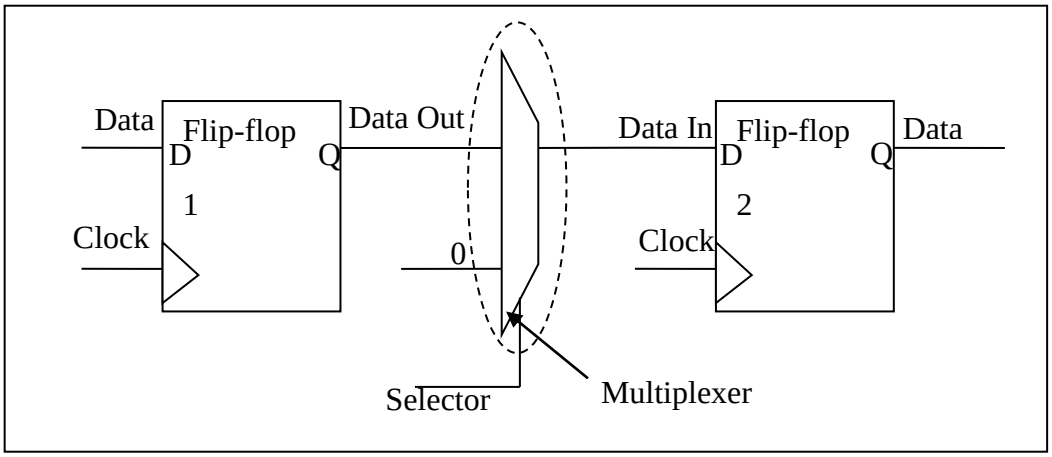

Figure 2.3: Data-gated flip-flop

Effective data-gating can reduce the short circuit power significantly. The typical data connection, the initial activity factor is rarely near 1. In theory, data-gating will provides less benefit than clock-gating in a typical design. For a high bus activity design like interconnect router, data-gating can provides better efficiency.

Dynamic voltage and frequency scaling (DVFS) becomes popular since it provides the best efficiency available technique on the market. DVFS is a technique where the operating frequency and the supply voltage can be dynamically adjusted bases on application usage. The effectiveness of DVFS technique highly depends on the

applications. This technique can be complimented with clock-gating which can give further power reduction. Referring to Equation 2.1, supply voltage gives the square effect on the dynamic power consumption. Table 2.2 shows the effectiveness of DVFS technique on the dynamic power consumption.

Table 2.2: Percentage Power Reduction with Different Frequency, Activity Factor and Voltage

P(delta)	η	feq (delta)	V (delta)
+0%	1	+0%	+0%
-50%	1	-50%	+0%
-43.75%	1	+0%	-25%
-57.81%	1	-25%	-25%
-71.88%	1	-50%	-25%
-68.36%	0.75	-25%	-25%

Intel® had Implemented DVFS technique into their microprocessors which named Enhanced Intel SpeedStep® Technology, EIST (Intel Corp., 2004). This technique works by down-clocking and under-voltage the processor while the system is in idle state. Another common DVFS application is the graphic processor unit (GPU). The GPU is used from very low workload such as to draw the image for display to very compute-intensive such as floating point calculation in 3 dimensions (3-D) graphic application. Table 2.3 shows a clock profile from NVIDIA® GeForce® series GPU (Nvidia Corp. 2010). The power consumption of the GPU with respective clock profile is also shown in figure 2. By reducing the clock and voltage of the GPU, the power consumption is significantly reduced.

Table 2.3 : Clock and Voltage profile for GTX 570 GPU from Nvidia®

	Core Clock	Memory Clock	GPU Voltage	Power Consumption
Desktop	51 MHz	68 MHz	0.91 V	25W
Blu-ray Playback	405 MHz	162 MHz	0.91 V	33W
3D Load	742 MHz	950 MHz	1.03 V	190W

2.3. Leakage Power Reduction Techniques

Leakage power becomes increasingly difficult to handle when the transistor size shrunk.

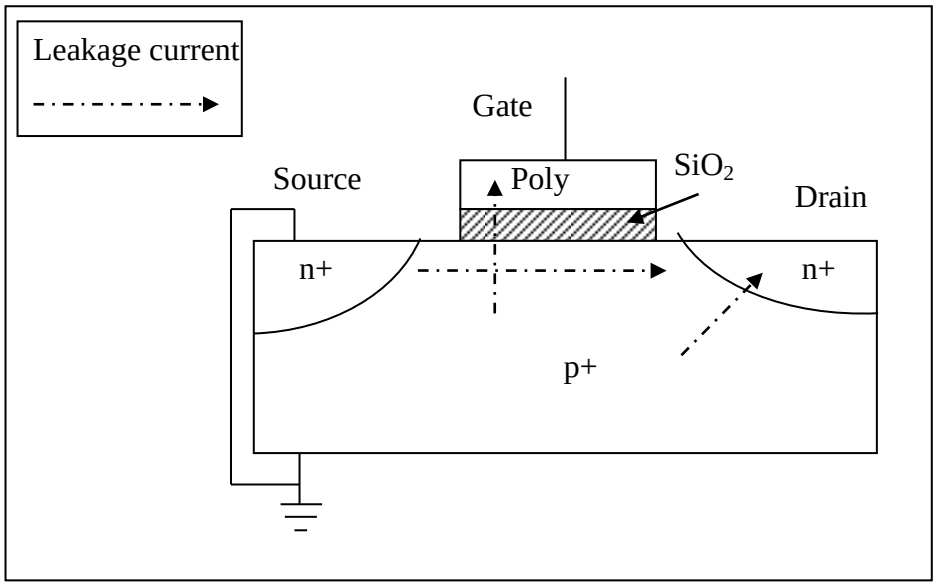

Figure 2.4: NMOS Transistor with Leakage Current

Figure 2.4 shows a bulk NMOS with three major leakage current sources. The three leakage currents are Junction leakage (from substrate to Drain), Channel leakage (from Source to Drain) and Gate tunneling leakage (from substrate to Gate).

To investigate the effectiveness of leakage power reduction techniques, the leakage power equation has to be analyzed deeper. The leakage power equation is given by:

$$P_l = VI_{leak} \text{ and } I_{leak} = i_s(e^{\frac{qV}{kT}} - 1) \quad [2.7]$$

where V = supply voltage, I_{leak} =leakage current, i_s = reverse saturation current, q = electronic charge (1.602×10^{-19} C), k =Boltzmann's constant ($1.38 \cdot 10^{-23}$ J/K), T=Temperature. From Equation 2.5, leakage power highly dependence of supply voltage and leakage current. Reducing supply voltage not only can reduce leakage power but also dynamic power consumption as discussed in Chapter 2.2. However, lower supply voltage can greatly impacted the timing of the design. Reverse saturation current is affected by threshold voltage of a CMOS. Higher temperature will also leads to higher leakage power consumption.

Power gating is the most common way to reduce leakage power by shutting off unused portion of the design. Since the power domain of the design needs to be properly define, this impact the design from architectural level to structural layout. Figure 2.5 shows a simple design with power gating technique implemented.

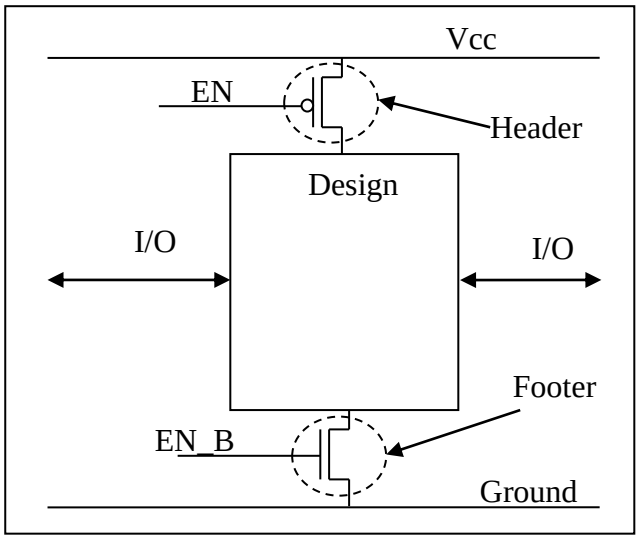

Figure 2.5: Power-gated design (Youngsoo Shin, 2010)

The design can be gated with Header or Footer depends on the design constraints. Footer provides faster turning between ON and OFF state with smaller area but with higher leakage current through the Footer. The Header will consume bigger area and turning off slower but excellent in low leakage current through Header. Power gating needs few extra components like isolation circuit and data retention for the design to works properly. Other concerns like power supply ripple during turning ON or OFF a cluster and time for the power supply to achieve proper voltage level will make the design much more complicated. From Figure 2.5, power gating can greatly reduce the supply voltage and leakage current.

Another way of reducing leakage power is multi-threshold voltage design. There are certain areas in the design which are timing critical. Examples are the memory, high-speed I/O, and execution unit. Some other parts are not timing-critical like interrupt handler, low-speed I/O, and legacy modules. By using different threshold voltages for different segments of the design can reduce the leakage power consumption without

significant performance penalty. High threshold voltage transistor switching slower but more resistant toward leakage current is suitable for low speed components while lower threshold voltage transistor used for timing-critical components to meet timing convergence profile. In order to implement this technique, a separate technology library is needed for both low and high threshold model. Multi-threshold voltage process will impose higher cost for fabrication and gives higher process variation in deep sub-micron design.

Figure 2.6 shows that in order to drive higher voltage logic from a lower voltage level, a DC-DC converter is required to connect them. The DC-DC converter itself will consumes certain amount of power, a further analysis is require to implement the multi-voltage technique to prevent advert effect on power consumption.

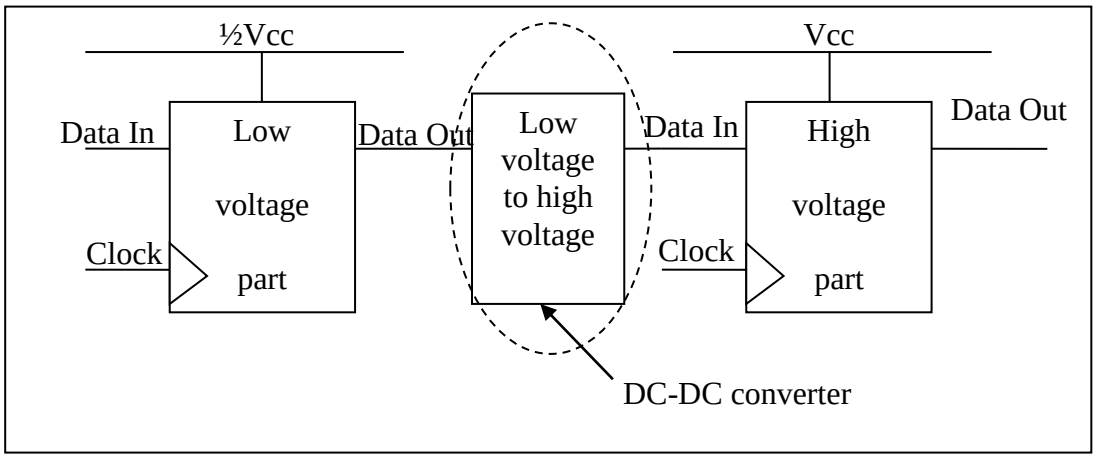

Figure 2.6: Two voltage level designs connected using DC-DC converter

Multi-supply voltage technique works the similar way as multi-threshold voltage technique except that the higher supply voltage was applied to timing-critical modules. However, due to differences in voltages, a DC-DC converter is needed to connect two parts in order for the design to works properly.

Substrate biasing is a more advance technique to reduce leakage power. By biasing the substrate, the threshold voltage can be dynamically adjusted. During heavy switching activities, low threshold voltage is desired to meet timing constraint while during idle mode; high threshold is desired to keep the leakage as low as possible.

2.4. Advance Power Convergence Technique

Most of the power convergence techniques are focusing on reducing power consumption while traded-off performance as consequences. In an extremely competitive market, performance is a relatively important criterion. Intel® has introduced a more advance power convergence technique in the Core™ i5 and Core™ i7 families of microprocessors named Turbo Boost Technology (Intel Corp., 2008). Turbo Boost Technology works by boosting up the operating frequency beyond the nominal frequency when detected a high workload while still works within the same thermal design power (TDP). Figure 2.7 shows operation frequencies of a quad-cores Core™ i7 microprocessor. When there is no active thread running, the processor will follows DVFS scheme to lower its power consumption. When a single heavily thread is detected, Turbo Boost kicked in to increase the operating frequency of an execution core to run beyond the nominal frequency while kept the remaining cores in low power state. As the number of active threads is increased, the extra bins provided by Turbo Boost reduced in order to keep the power consumption and temperature within the TDP.

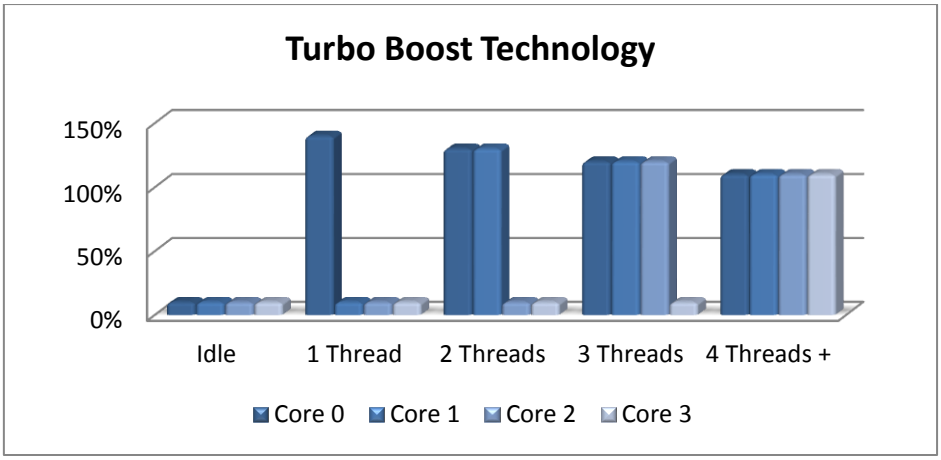

Figure 2.7: Turbo Boost Technology allows the operation frequency to run beyond the nominal frequency

Advanced Micro Device (AMD) has implemented a similar but less efficient technology named AMD Turbo CORE Technology in hexa-cores AMD Phenom™ II X6 microprocessor (AMD Inc., 2011). AMD Turbo CORE Technology only works in 2 modes, 3-cores boost or 6-cores boost. When a single thread is detected, 3 out of the 6 cores are bumped up while the remaining 3 cores are kept in low power state. If multiple threads are detected, all 6 cores are boosted up with fewer bins.

2.5. Other Power Related Factors

2.5.1. System Application and Software

System and Software applications are one of the major contributions of power usage efficiency. One of the good examples is the proper hardware usage in the system. As discussed in previous section, a GPU is typically used for graphic related tasks. Using a general purpose processer to work in graphic related tasks not only the performance is lagging, much more power are needed to run the same task compare to a GPU which is

design specifically for the graphic related task. Software also plays a major role in this area. In order for an operating system (OS) to work together with hardware, a good driver is needed for proper protocol communication. An excellent driver can make the hardware works as efficient as possible. Same condition happened in the server area. For a typical server or a data centre, highly parallel threads are more important than a high operating frequency.

2.5.2. Interconnects and Devices (Transistor)

Apart from the IC design engineers' effort to optimize the efficiency of the IC, chemical engineering and physic engineering are also searching multiple ways to reduce the power consumption of the IC device in molecular level. One of the migrations is the changes of material used for interconnects of the transistors. Changing the material from Aluminum to Copper greatly improve the voltage drop across the power supply interconnects in the IC which leads to lower power loss during the power transmission. Other than the copper, other materials are being observed to replace soon to be aging copper as the transistor shrinking process continues. Optical transmission in-a-chip is currently being undergoes research by IBM® to remove the losses of the interconnects by using ray light rather than electrical signal (IBM, 2012). The technology is still in infant state to be commercialized.

On the device side, other than shrinking the transistor size, other efforts are being done to bring down the total power consumption of the IC. High-K dielectric Metal Gate (HKMG) is one of the process technologies which can helps to reduce the gate leakage

of the transistor (Intel Corp., 2003). Another approach is Silicon on Insulator (SOI) bulk process technology which targeted to reduce the substrate leakage (Vishwas Jaju, 2004). Intel® announced a major breakthrough and historic innovation in microchips with the world's first 3-D tri-gate transistors in mass production (Intel Corp. 2011). This process technology allow a much better control on the leakage power which make the shrinking process of the transistor can proceed further.

During the past, custom layout was the way to implement the design from the schematic. By during custom layout, the main constraints of the design can be optimized. The timing, area and power are the key constraints that being optimized by combining a few transistors common wells. For older process technology, the transistor's size is larger than the interconnect. By shrinking the transistor substrate to smaller size, the dynamic power can be reduced and the critical path can be improved. When the transistor's size is entering the deep sub-micron region, the benefits of custom layout design starts to diminish. The tedious work on the custom layout is not a practical work flow and hence the introducing of standard cells library. Figure 2.8 shows the standard cell in 250nm.

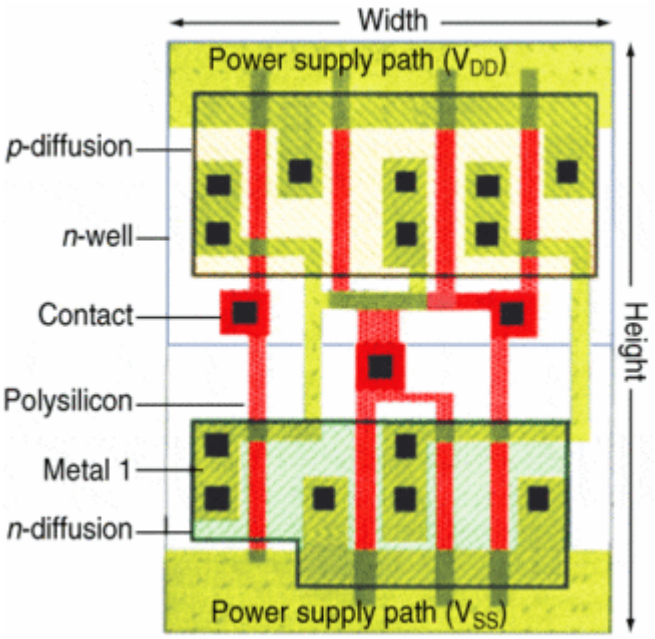

Figure 2.8: Sample of a Standard Cell (Jun Wang and Alfred K. Wong, 2001)

Standard cells library has the pre-optimized transistors into a block of cells with standard height. The design engineers can then optimize the layout during placement and routing. By doing so, the power optimization be done in a more efficient way.

There are several discoveries on the replacement of the typical transistor model. The most renowned technology is the Carbon Nanotube technology (Adrian Bachtold, 2001). With properly align carbon molecules in a tube form, a single electron can be used to transfer the signal from one to the other. This can reduce the power consumption to absolute minimum in a device perspective. However, the process of creating Carbon Nanotube is generally far from mass production capability in current stage.